GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 30, 2002, 10:20:38 ; Search time 35.8079 Seconds
(without alignments)
2307.182 Million cell updates/sec

Title: US-10-054-680-4
Perfect score: 3228
Sequence: 1 MAWLRLQPLTSAFLHFGLVT.....ADYGRRGGQEDSRDGKASIG 620

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3093	95.8	921	23	ABB83246	Human transporter
2	3093	95.8	927	23	AAM47745	Human natrium(+)-c
3	3016	93.4	927	23	ABB83247	Human transporter
4	2147.5	66.5	970	23	AAE18291	Bovine NCX-1 prote
5	2143.5	66.4	609	22	ABB32633	Peptide #139 encod
6	2143.5	66.4	609	22	ABB18131	Protein #130 encod
7	2143.5	66.4	609	22	AAM53461	Human brain expres
8	2143.5	66.4	609	22	AAM13701	Peptide #135 encod
9	2143.5	66.4	609	22	AAM26102	Peptide #139 encod
10	2143.5	66.4	609	23	ABG35474	Human peptide enco
11	1961	60.7	952	21	AAB41497	Human ORFX ORF1261
12	1339.5	41.5	950	22	ABB61721	Drosophila melanog
13	362.5	11.2	539	21	AAY58044	Arabidopsis thalia
14	203.5	6.3	618	22	AAM78712	Human protein SEQ
15	197	6.1	546	22	ABB12262	Human very large G
16	180.5	5.6	603	23	AAE22088	Human novel ion ex
17	171.5	5.3	2071	22	AAE10949	Mouse mass1 protei
18	171.5	5.3	2780	22	AAE10924	Mouse monogenic au
19	162.5	5.0	856	22	ABB60425	Drosophila melanog
20	161	5.0	759	23	ABG61545	Human transporter
21	158.5	4.9	1615	23	ABB05663	Human signal trans
22	154.5	4.8	658	22	ABB57820	Drosophila melanog
23	154	4.8	2777	22	AAE10925	Human monogenic au
24	146	4.5	35	22	AAE10945	Mouse Na+/Ca2+ exc
25	130.5	4.0	1451	22	AAB95868	Human protein sequ
26	130.5	4.0	3530	22	AAM25586	Human protein sequ
27	126.5	3.9	1822	15	AAR55273	Beta subunit of in
28	126.5	3.9	1873	23	AAE14708	Human beta4 integr
29	126.5	3.9	1873	23	AAE14714	Human beta4 integr
30	126.5	3.9	1875	22	AAB68089	Amino acid sequenc
31	126.5	3.9	3105	23	ABB80604	Human sbg1018172CS
32	124	3.8	570	21	AAG36431	Arabidopsis thalia
33	116.5	3.6	584	22	AAE04892	Human transporter
34	116	3.6	463	23	AAE21674	Chicken DP214-like
35	114	3.5	1466	23	AAE14709	Mouse beta4 integr
36	113.5	3.5	339	23	ABP38763	Staphylococcus epi
37	113.5	3.5	353	22	ABB58318	Drosophila melanog
38	109	3.4	1630	22	ABB59874	Drosophila melanog
39	108.5	3.4	652	15	AAR63808	Human amphotropic
40	108.5	3.4	652	19	AAW70499	Human sodium-lithi
41	107.5	3.3	494	22	AAU56619	Propionibacterium
42	107.5	3.3	918	22	ABB59819	Drosophila melanog
43	107	3.3	431	23	ABP27034	Streptococcus poly
44	106	3.3	646	21	AAB18188	Plasmodium falcipa
45	105	3.3	404	20	AAY35300	Chlamydia pneumoni

ALIGNMENTS

RESULT 1
ABB83246
ID ABB83246 standard; Protein; 921 AA.
XX
AC ABB83246;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human transporter protein.
XX
KW Human; sodium/calcium exchanger; transporter; brain; heart; kidney; lung;
KW spleen; testis; leukocyte; foetal brain; chromosome 14.
XX
OS Homo sapiens.
XX
PN WO200233086-A2.
XX
PD 25-APR-2002.
XX
PF 17-OCT-2001; 2001WO-US32152.
XX
PR 17-OCT-2000; 2000US-240836P.
PR 13-MAR-2001; 2001US-0804474.
XX
PA (PEKE) PE CORP NY.
XX
PI Merkulov GV, Ketchum KA, Shao W, Yan C, Di Francesco V;
PI Beasley EM;
XX
DR WPI; 2002-479677/51.
DR N-PSDB; ABN83428, ABN83429.
XX
PT Human transporter peptide related to sodium/calcium exchanger subfamily

PT for identifying modulators useful for treating a disease or condition
PT mediated by human transporter protein -
XX
PS Claim 1; Fig 2; 200pp; English.
XX
CC The present sequence is a human transporter protein, which is related to
CC the sodium/calcium exchanger subfamily. Experimental data indicates
CC expression of the transporter gene in humans in brain, heart, kidney,
CC lung, spleen, testis, leukocyte and foetal brain. The gene of the
CC transporter was mapped to chromosome 14 by ePCR.
XX
SQ Sequence 921 AA;

Query Match 95.8%; Score 3093; DB 23; Length 921;
Best Local Similarity 100.0%; Pred. No. 3.1e-306;
Matches 595; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVIL 60
Db 1 MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVIL 60

Qy 61 PIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGE 120
Db 61 PIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGE 120

Qy 121 TSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFI 180
Db 121 TSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFI 180

Qy 181 IIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFF 240
Db 181 IIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFF 240

Qy 241 FPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDGN 300
Db 241 FPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDGN 300

Qy 301 LVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATR 360
Db 301 LVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATR 360

Qy 361 MMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLTVVR 420
Db 361 MMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLTVVR 420

Qy 421 KGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHF 480
Db 421 KGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHF 480

Qy 481 FVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIH 540
Db 481 FVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIH 540

Qy 541 VSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETV 595
Db 541 VSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETV 595

RESULT 2
AAM47745
ID AAM47745 standard; Protein; 927 AA.
XX
AC AAM47745;
XX
DT 25-FEB-2002 (first entry)
XX
DE Human natrium(+)-calcium(2+) exchanger form 3 protein, HNCX3.
XX
KW Human; Natrium(+)-Calcium(2+) exchanger form 3; HNCX3; chromosome 14;
KW cardiac failure; myocardial infarction; cardiac hypertrophy; arrhythmia;
KW myocarditis; pulmonary hypertension; cardiotoxicity; cardiant; Vaccine;
KW coronary heart disease; renal failure; ischaemic disorder;
KW Antiarrhythmic; Vasotropic; Hypotensive; cardiovascular disorder.
XX
OS Homo sapiens.
XX
PN WO200183744-A2.
XX
PD 08-NOV-2001.
XX
PF 30-APR-2001; 2001WO-EP04886.
XX
PR 02-MAY-2000; 2000EP-0109080.
XX
PA (MERE) MERCK PATENT GMBH.
XX
PI Wilm C;
XX
DR WPI; 2002-041493/05.
DR N-PSDB; ABA04756.
XX
PT New polypeptide, useful as vaccines for inducing immune response
PT against diseases such as myocardial infarction, arrhythmia, ischemic
PT disorders, renal disorders in mammal -
XX
PS Claim 1; Page 38-41; 41pp; English.
XX
CC The present sequence is the protein sequence for human Natrium(+)-Calcium
CC (2+) exchanger form 3 (HNCX3). The HNCX3 gene maps to human chromosome
CC 14. HNCX3 and its coding sequence are useful for treating acute and
CC chronic cardiac failure of different aetiologies, myocardial infarction,
CC cardiac hypertrophy, arrhythmia, myocarditis, pulmonary hypertension,
CC cardiotoxicity (e.g. induced by chemotherapy), coronary heart disease,
CC acute and chronic renal failure, ischaemic disorders of skeletal muscle
CC and ischaemic brain disorders of different aetiologies.
XX
SQ Sequence 927 AA;

Query Match 95.8%; Score 3093; DB 23; Length 927;
Best Local Similarity 100.0%; Pred. No. 3.2e-306;
Matches 595; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVIL 60
Db 1 MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVIL 60

Qy 61 PIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGE 120
Db 61 PIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGE 120

Qy 121 TSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFI 180
Db 121 TSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFI 180

Qy 181 IIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFF 240
Db 181 IIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFF 240

Qy 241 FPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDGN 300
Db 241 FPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDGN 300

Qy 301 LVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATR 360
Db 301 LVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATR 360

Qy 361 MMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLTVVR 420
Db 361 MMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLTVVR 420

Qy 421 KGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHF 480
Db 421 KGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHF 480

Qy 481 FVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIH 540

```
Db    481 FVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIH 540

Qy    541 VSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETV 595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 VSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETV 595


RESULT 3
ABB83247
ID   ABB83247 standard; Protein; 927 AA.
XX
AC   ABB83247;
XX
DT   21-AUG-2002  (first entry)
XX
DE   Human transporter protein-related protein, used in a homology alignment.
XX
KW   Human; sodium/calcium exchanger; transporter; brain; heart; kidney; lung;
KW   spleen; testis; leukocyte; foetal brain; chromosome 14.
XX
OS   Unidentified.
XX
PN   WO200233086-A2.
XX
PD   25-APR-2002.
XX
PF   17-OCT-2001; 2001WO-US32152.
XX
PR   17-OCT-2000; 2000US-240836P.
PR   13-MAR-2001; 2001US-0804474.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Merkulov GV,  Ketchum KA,  Shao W,  Yan C,  Di Francesco V;
PI   Beasley EM;
XX
DR   WPI; 2002-479677/51.
XX
PT   Human transporter peptide related to sodium/calcium exchanger subfamily
PT   for identifying modulators useful for treating a disease or condition
PT   mediated by human transporter protein  -
XX
PS   Disclosure; Fig 2; 200pp; English.
XX
CC   The present invention relates to a human transporter protein, which is
CC   related to the sodium/calcium exchanger subfamily (ABB83246).
CC   Experimental data indicates expression of the transporter gene in humans
CC   in brain, heart, kidney, lung, spleen, testis, leukocyte and foetal
CC   brain. The gene of the transporter was mapped to chromosome 14 by ePCR.
CC   The present protein was used in a sequence alignment with the transporter
CC   protein to illustrate the invention.
XX
SQ   Sequence   927 AA;

  Query Match             93.4%;  Score 3016;  DB 23;  Length 927;
  Best Local Similarity   97.1%;  Pred. No. 2.3e-298;
  Matches  578;  Conservative    8;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGVIL 60
          ||||||||||||||||||||||||||||||||    |||| ||||||||||||||||||
Db      1 MAWLRLQPLTSAFLHFGLVTFVLFLNGLRAEAGDLRDVPSAGQNNESCSGSSDCKEGVIL 60

Qy     61 PIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGE 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGE 120

Qy    121 TSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFI 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNMFI 180

Qy    181 IIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLFF 240
          ||||||||||||||||||||||:|||||:||||||||||||||||||||||||||||||||
Db    181 IIGICVYVIPDGETRKIKHLRVFFVTAAWSVFAYIWLYMILAVFSPGVVQVWEGLLTLFF 240

Qy    241 FPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDGN 300
          ||||||||||||||||||||||:|||||||||||||:||||||||||||||||||||||||
Db    241 FPVCVLLAWVADKRLLFYKYMHKRYRTDKHRGIIIETEGEHPKGIEMDGKMMNSHFLDGN 300

Qy    301 LVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATR 360
          |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 LIPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATR 360

Qy    361 MMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGAVLLTVVR 420
          |||||||||||||||||| :|||||||||||| ||||||||||||||||||||||||||||
Db    361 MMTGAGNILKKHAAEQAKKTASMSEVHTDEPEDFASKVFFDPCSYQCLENCGAVLLTVVR 420

Qy    421 KGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHF 480
          ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 KGGDISKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHF 480

Qy    481 FVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIH 540
          ||||||||:|||| |||| ||| ||||||||||||||||||||||||||||||||||||||
Db    481 FVRLSNVRVEEEQLEEGMTPAILNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTIH 540

Qy    541 VSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETV 595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 VSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGELEFKNDETV 595


RESULT 4
AAE18291
ID   AAE18291 standard; Protein; 970 AA.
XX
AC   AAE18291;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Bovine NCX-1 protein.
XX
KW   Bovine; recombinant protein; larvae expression system; membrane protein;
KW   transport protein; cardiac sodium-calcium exchange protein; Na-K ATPase;
KW   NCX1; cystic fibrosis transmembrane conductance regulator; CFTR; vaccine;
KW   channel forming protein; junctional protein; conexin 32.
XX
OS   Bos taurus.
XX
PN   WO200206464-A2.
XX
PD   24-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-US21606.
XX
PR   13-JUL-2000; 2000US-218125P.
XX
PA   (UMOR ) UNIV MISSOURI.
XX
PI   Hale CC,  Price EM;
XX
DR   WPI; 2002-171806/22.
DR   N-PSDB; AAD24450.
XX
PT   Producing recombinant proteins e.g. membrane, transport and channel
PT   forming proteins in larvae expression system, by infecting larvae with
PT   vector having a sequence encoding recombinant fusion protein with
PT   affinity tag  -
XX
PS   Example 1; Page 37-40; 40pp; English.
XX
CC   The patent discloses methods of producing recombinant proteins in larvae
CC   expression system, by infecting the larvae with vector having a sequence
CC   encoding recombinant fusion protein with affinity tag. The methods are
CC   useful for producing recombinant protein, preferably membrane proteins,
CC   transport proteins such as NCX1 (cardiac sodium-calcium exchange protein)
CC   or Na-K ATPase, channel forming proteins such as cystic fibrosis trans-
CC   membrane conductance regulator (CFTR), junctional protein (conexin 32),
CC   receptor, cytoskeletal and other membrane associated proteins. They are
```

```
CC   also useful for producing prostate specific membrane antigens and sodium
CC   phosphate co-transporters from kidney. The methods are also useful for
CC   producing recombinant fusion proteins in large quantities that are both
CC   highly homogenous and biologically active. The recombinant proteins
CC   produced by the methods of the invention can be included as part of a
CC   pharmaceutical, nutritional, drug or vaccine composition. The present
CC   sequence is bovine NCX-1 protein.
XX
SQ   Sequence   970 AA;

  Query Match             66.5%;  Score 2147.5;  DB 23;  Length 970;
  Best Local Similarity   69.4%;  Pred. No. 1.5e-209;
  Matches  422;  Conservative   74;  Mismatches   91;  Indels   21;  Gaps    8;

Qy       1 MAWLRLQPLTSAFLHFGLVTFVLF--LNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV 58
Db       1 MLQFSLSPTLSMGFHVIAMVALLFSHVDHISAETEMEGEGNETGE----CTGSYYCKKGV 56

Qy      59 ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPN 118
Db      57 ILPIWEPQDPSFGDKIARATVYFVAMVYMFLGVSIIADRFMSSIEVITSQEKEITIKKPN 116

Qy     119 GETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNM 178
Db     117 GETTKTTVRIWNETVSNLTLMALGSSAPEILLSVIEVCGHNFTAGDLGPSTIVGSAAFNM 176

Qy     179 FIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTL 238
Db     177 FIIIALCVYVVPDGETRKIKHLRVFFVTAAWSIFAYTWLYIILSVSSPGVVEVWEGLLTF 236

Qy     239 FFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKG---IEMDGKMMNSH 295
Db     237 FFFPICVVFAWVADRRLLFYKYVYKRYRAGKQRGMIIEHEGDRPSSKTEIEMDGKVVNSH 296

Qy     296 ---FLDGNLVPLEGKEVD----ESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ 348
Db     297 VDSFLDGALV-LEVDERDQDDEEARREMARILKELKQKHPEKEIEQLIELANYQVLSQQQ 355

Qy     349 KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPE-DFISKVFFDPCSYQC 407
Db     356 KSRAFYRIQATRLMTGAGNILKRHAADQARKAVSMHEVNTEVAENDPVSKIFFEQGTYQC 415

Qy     408 LENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVG 467
Db     416 LENCGTVALTIIRRGGDLTNTVFVDFRTEDGTANAGSDYEFTEGTVVFKPGETQKEIRVG 475

Qy     468 IIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDD 527
Db     476 IIDDDIFEEDENFLVHLSNVKVSLEASEDGILEASHVS---TLACLGSPSTATVTIFDDD 532

Qy     528 HAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGEL 587
Db     533 HAGIFTFEEPVTHVSESIGIMEVKVLRTSGARGNVIVPYKTIEGTARGGGEDFEDTCGEL 592

Qy     588 EFKNDETV 595
Db     593 EFQNDEIV 600


RESULT 5
ABB32633
ID   ABB32633 standard; Peptide; 609 AA.
XX
AC   ABB32633;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #139 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 27; SEQ ID NO 25268; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a peptide encoded by a single exon
CC   nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   609 AA;

  Query Match             66.4%;  Score 2143.5;  DB 22;  Length 609;
  Best Local Similarity   69.3%;  Pred. No. 1.8e-209;
  Matches  420;  Conservative   76;  Mismatches   89;  Indels   21;  Gaps    8;

Qy       1 MAWLRLQPLTSAFLHFGLVTFVLF--LNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV 58
Db      11 MRRLSLSPTFSMGFHLLVTVSLLFSHVDHVIAETEMEGEGNETGE----CTGSYYCKKGV 66

Qy      59 ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPN 118
Db      67 ILPIWEPQDPSFGDKIARATVYFVAMVYMFLGVSIIADRFMSSIEVITSQEKEITIKKPN 126

Qy     119 GETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNM 178
Db     127 GETTKTTVRIWNETVSNLTLMALGSSAPEILLSVIEVCGHNFTAGDLGPSTIVGSAAFNM 186

Qy     179 FIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTL 238
Db     187 FIIIALCVYVVPDGETRKIKHLRVFFVTAAWSIFAYTWLYIILSVISPGVVEVWEGLLTF 246

Qy     239 FFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKG---IEMDGKMMNSH 295
Db     247 FFFPICVVFAWVADRRLLFYKYVYKRYRAGKQRGMIIEHEGDRPSSKTEIEMDGKVVNSH 306

Qy     296 ---FLDGNLVPLEGKEVD----ESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ 348
Db     307 VENFLDGALV-LEVDERDQDDEEARREMARILKELKQKHPDKEIEQLIELANYQVLSQQQ 365

Qy     349 KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPE-DFISKVFFDPCSYQC 407
Db     366 KSRAFYRIQATRLMTGAGNILKRHAADQARKAVSMHEVNTEVTENDPVSKIFFEQGTYQC 425

Qy     408 LENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVG 467
Db     426 LENCGTVALTIIRRGGDLTNTVFVDFRTEDGTANAGSDYEFTEGTVVFKPGDTQKEIRVG 485
```

```
Qy    468 IIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDD 527
          ||||||||||||:| | ||||::  |  |:|:  |   |     | | ||  ||||| |||
Db    486 IIDDDIFEEDENFLVHLSNVKVSSEASEDGILEANHVS---TLACLGSPSTATVTIFDDD 542

Qy    528 HAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGEL 587
          ||||||||    |||||||:||||||||||||| ||||::|:||||:||||||||| |||
Db    543 HAGIFTFEEPVTHVSESIGIMEVKVLRTSGARGNVIVPYKTIEGTARGGGEDFEDTCGEL 602

Qy    588 EFKNDE 593
          ||:|||
Db    603 EFQNDE 608


RESULT 6
ABB18131
ID   ABB18131 standard; Protein; 609 AA.
XX
AC   ABB18131;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Protein #130 encoded by probe for measuring heart cell gene expression.
XX
KW   Human; gene expression; heart; microarray; vascular system;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
 X
 R   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 15; SEQ ID No 19901; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart (see
CC   ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC   probe. The probes may be used for predicting, measuring and displaying
CC   gene expression in samples derived from the human heart via microarrays.
CC   By measuring gene expression, the probes are useful for predicting,
CC   diagnosing, grading, staging, monitoring and prognosing diseases of the
CC   human heart and vascular system e.g. cardiovascular disease,
CC   hypertension, cardiac arrhythmias and congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   609 AA;

  Query Match           66.4%;  Score 2143.5;  DB 22;  Length 609;
  Best Local Similarity  69.3%;  Pred. No. 1.8e-209;
  Matches  420;  Conservative   76;  Mismatches   89;  Indels   21;  Gaps    8;

Qy      1 MAWLRLQPLTSAFLHFGLVTFVLF--LNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV 58
          |  | | |  |   |  :  :||  :: : ||    |:   ||:    |:|| ||:||
Db     11 MRRLSLSPTFSMGFHLLVTVSLLFSHVDHVIAETEMEGEGNETGE----CTGSYYCKKGV 66

Qy     59 ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPN 118
          ||||| |::|| ||||||  |||||::|||||||||||||:|||||||||:|:||||||
Db     67 ILPIWEPQDPSFGDKIARATVYFVAMVYMFLGVSIIADRFMSSIEVITSQEKEITIKKPN 126

Qy    119 GETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNM 178
          |||: ||:|:|||||||||||||||||||||||:|||||| | ||||||||||||||||
Db    127 GETTKTTVRIWNETVSNLTLMALGSSAPEILLSVIEVCGHNFTAGDLGPSTIVGSAAFNM 186

Qy    179 FIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTL 238
          |||| :||||:|||||||||||||:|||||||| |||:||:| |||||:|||||||
Db    187 FIIIALCVYVVPDGETRKIKHLRVFFVTAAWSIFAYTWLYIILSVISPGVVEVWEGLLTF 246

Qy    239 FFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKG---IEMDGKMMNSH 295
          ||||:||: |||||:|||||||::|:||  | ||:||| ||| |     ||||||::|||
Db    247 FFFPICVVFAWVADRRLLFYKYVYKRYRAGKQRGMIIEHEGDRPSSKTEIEMDGKVVNSH 306

Qy    296 ---FLDGNLVPLEGKEVD----ESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ 348
             |||| || ||  | |    |:|||| ||||:||||||:|:::||:|:|||  || ||
Db    307 VENFLDGALV-LEVDERDQDDEEARREMARILKELKQKHPDKEIEQLIELANYQVLSQQQ 365

Qy    349 KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPE-DFISKVFFDPCSYQC 407
          ||||||||||:|||||||:|||:||:|| || ||:|:  | | :||:||:  :|||
Db    366 KSRAFYRIQATRLMTGAGNILKRHAADQARKAVSMHEVNTEVTENDPVSKIFFEQGTYQC 425

Qy    408 LENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVG 467
          ||||| | ||::|:|||:: |::||::||||:||||:|||||||||| |||:|||| ||
Db    426 LENCGTVALTIIRRGGDLTNTVFVDFRTEDGTANAGSDYEFTEGTVVFKPGDTQKEIRVG 485

Qy    468 IIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDD 527
          ||||||||||||:| | ||||::  |  |:|:  |   |     | | ||  ||||| |||
Db    486 IIDDDIFEEDENFLVHLSNVKVSSEASEDGILEANHVS---TLACLGSPSTATVTIFDDD 542

Qy    528 HAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGEL 587
          ||||||||    |||||||:||||||||||||| ||||::|:||||:||||||||| |||
Db    543 HAGIFTFEEPVTHVSESIGIMEVKVLRTSGARGNVIVPYKTIEGTARGGGEDFEDTCGEL 602

Qy    588 EFKNDE 593
          ||:|||
Db    603 EFQNDE 608


RESULT 7
AAM53461
ID   AAM53461 standard; Protein; 609 AA.
XX
AC   AAM53461;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 25566.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
```

```
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 25566; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention.
XX
     Sequence   609 AA;

  Query Match             66.4%;  Score 2143.5;  DB 22;  Length 609;
  Best Local Similarity   69.3%;  Pred. No. 1.8e-209;
  Matches  420;  Conservative   76;  Mismatches   89;  Indels   21;  Gaps    8;

Qy       1 MAWLRLQPLTSAFLHFGLVTFVLF--LNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV 58
Db      11 MRRLSLSPTFSMGFHLLVTVSLLFSHVDHVIAETEMEGEGNETGE----CTGSYYCKKGV 66

Qy      59 ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPN 118
Db      67 ILPIWEPQDPSFGDKIARATVYFVAMVYMFLGVSIIADRFMSSIEVITSQEKEITIKKPN 126

Qy     119 GETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNM 178
Db     127 GETTKTTVRIWNETVSNLTLMALGSSAPEILLSVIEVCGHNFTAGDLGPSTIVGSAAFNM 186

Qy     179 FIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTL 238
Db     187 FIIIALCVYVVPDGETRKIKHLRVFFVTAAWSIFAYTWLYIILSVISPGVVEVWEGLLTF 246

Qy     239 FFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKG---IEMDGKMMNSH 295
Db     247 FFFPICVVFAWVADRRLLFYKYVYKRYRAGKQRGMIIEHEGDRPSSKTEIEMDGKVVNSH 306

Qy     296 ---FLDGNLVPLEGKEVD----ESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ 348
Db     307 VENFLDGALV-LEVDERDQDDEEARREMARILKELKQKHPDKEIEQLIELANYQVLSQQQ 365

Qy     349 KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPE-DFISKVFFDPCSYQC 407
Db     366 KSRAFYRIQATRLMTGAGNILKRHAADQARKAVSMHEVNTEVTENDPVSKIFFEQGTYQC 425

Qy     408 LENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVG 467
Db     426 LENCGTVALTIIRRGGDLTNTVFVDFRTEDGTANAGSDYEFTEGTVVFKPGDTQKEIRVG 485

Qy     468 IIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDD 527
Db     486 IIDDDIFEEDENFLVHLSNVKVSSEASEDGILEANHVS---TLACLGSPSTATVTIFDDD 542

Qy     528 HAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGEL 587
Db     543 HAGIFTFEEPVTHVSESIGIMEVKVLRTSGARGNVIVPYKTIEGTARGGGEDFEDTCGEL 602

Qy     588 EFKNDE 593
Db     603 EFQNDE 608


RESULT 8
```

```
AAM13701
ID   AAM13701 standard; Protein; 609 AA.
XX
AC   AAM13701;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #135 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27; SEQ ID No 18527; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   609 AA;

  Query Match             66.4%;  Score 2143.5;  DB 22;  Length 609;
  Best Local Similarity   69.3%;  Pred. No. 1.8e-209;
  Matches  420;  Conservative   76;  Mismatches   89;  Indels   21;  Gaps    8;

Qy       1 MAWLRLQPLTSAFLHFGLVTFVLF--LNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV 58
Db      11 MRRLSLSPTFSMGFHLLVTVSLLFSHVDHVIAETEMEGEGNETGE----CTGSYYCKKGV 66

Qy      59 ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPN 118
Db      67 ILPIWEPQDPSFGDKIARATVYFVAMVYMFLGVSIIADRFMSSIEVITSQEKEITIKKPN 126

Qy     119 GETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNM 178
Db     127 GETTKTTVRIWNETVSNLTLMALGSSAPEILLSVIEVCGHNFTAGDLGPSTIVGSAAFNM 186

Qy     179 FIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTL 238
Db     187 FIIIALCVYVVPDGETRKIKHLRVFFVTAAWSIFAYTWLYIILSVISPGVVEVWEGLLTF 246
```

```
Qy    239 FFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKG---IEMDGKMMNSH 295
          ||||:||: |||||:|||||||::|:|| | ||:||| ||| |     ||||||::|||
Db    247 FFFPICVVFAWVADRRLLFYKYVYKRYRAGKQRGMIIEHEGDRPSSKTEIEMDGKVVNSH 306

Qy    296 ---FLDGNLVPLEGKEVD----ESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ 348
             |||| || ||  | |    |:|||| ||||:||||||:|:::||:|:||| || ||
Db    307 VENFLDGALV-LEVDERDQDDEEARREMARILKELKQKHPDKEIEQLIELANYQVLSQQQ 365

Qy    349 KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPE-DFISKVFFDPCSYQC 407
          |||||||||||||:|||||||||:|||:||:|| || ||:|:  | | :||:||:  :|||
Db    366 KSRAFYRIQATRLMTGAGNILKRHAADQARKAVSMHEVNTEVTENDPVSKIFFEQGTYQC 425

Qy    408 LENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVG 467
          ||||| | ||::|:|||:: |::||::||||:||||:|||||||||| |||:|||  ||
Db    426 LENCGTVALTIIRRGGDLTNTVFVDFRTEDGTANAGSDYEFTEGTVVFKPGDTQKEIRVG 485

Qy    468 IIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDD 527
          |||||||||||:| | ||||::  |  |:|:  |   |      | | ||  ||||| |||
 b    486 IIDDDIFEEDENFLVHLSNVKVSSEASEDGILEANHVS---TLACLGSPSTATVTIFDDD 542

Qy    528 HAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGEL 587
          ||||||||    |||||||:|||||||||||| ||||::|:||||:||||||||| |||
Db    543 HAGIFTFEEPVTHVSESIGIMEVKVLRTSGARGNVIVPYKTIEGTARGGGEDFEDTCGEL 602

Qy    588 EFKNDE 593
          ||:|||
Db    603 EFQNDE 608


RESULT 9
AAM26102
ID   AAM26102 standard; Protein; 609 AA.
XX
AC   AAM26102;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #139 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID No 26371; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   609 AA;

  Query Match             66.4%;  Score 2143.5;  DB 22;  Length 609;
  Best Local Similarity   69.3%;  Pred. No. 1.8e-209;
  Matches  420;  Conservative   76;  Mismatches   89;  Indels   21;  Gaps    8;

Qy      1 MAWLRLQPLTSAFLHFGLVTFVLF--LNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV 58
          |  | | |  |   |  :   :||  :: : ||    |:   ||:    |:||  ||:||
Db     11 MRRLSLSPTFSMGFHLLVTVSLLFSHVDHVIAETEMEGEGNETGE----CTGSYYCKKGV 66

Qy     59 ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPN 118
          ||||| |::|| ||||||  |||||::|||||||||||||:|||||||:|:||||||
Db     67 ILPIWEPQDPSFGDKIARATVYFVAMVYMFLGVSIIADRFMSSIEVITSQEKEITIKKPN 126

Qy    119 GETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNM 178
          |||: ||:|:|||||||||||||||||||||:|||||| | ||||||||||||||||||
Db    127 GETTKTTVRIWNETVSNLTLMALGSSAPEILLSVIEVCGHNFTAGDLGPSTIVGSAAFNM 186

Qy    179 FIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTL 238
          |||| :||||:|||||||||||||:|||||||| |||:||:| |||||:|||||||
Db    187 FIIIALCVYVVPDGETRKIKHLRVFFVTAAWSIFAYTWLYIILSVISPGVVEVWEGLLTF 246

Qy    239 FFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKG---IEMDGKMMNSH 295
          ||||:||: |||||:|||||||::|:||  | ||:||| ||| |     ||||||::|||
Db    247 FFFPICVVFAWVADRRLLFYKYVYKRYRAGKQRGMIIEHEGDRPSSKTEIEMDGKVVNSH 306

Qy    296 ---FLDGNLVPLEGKEVD----ESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ 348
             |||| || ||  | |    |:|||| ||||:||||||:|:::||:|:||| || ||
Db    307 VENFLDGALV-LEVDERDQDDEEARREMARILKELKQKHPDKEIEQLIELANYQVLSQQQ 365

Qy    349 KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPE-DFISKVFFDPCSYQC 407
          |||||||||||||:|||||||||:|||:||:|| || ||:|:  | | :||:||:  :|||
Db    366 KSRAFYRIQATRLMTGAGNILKRHAADQARKAVSMHEVNTEVTENDPVSKIFFEQGTYQC 425

Qy    408 LENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVG 467
          ||||| | ||::|:|||:: |::||::||||:||||:|||||||||| |||:|||  ||
Db    426 LENCGTVALTIIRRGGDLTNTVFVDFRTEDGTANAGSDYEFTEGTVVFKPGDTQKEIRVG 485

Qy    468 IIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDD 527
          |||||||||||:| | ||||::  |  |:|:  |   |      | | ||  ||||| |||
Db    486 IIDDDIFEEDENFLVHLSNVKVSSEASEDGILEANHVS---TLACLGSPSTATVTIFDDD 542

Qy    528 HAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGEL 587
          ||||||||    |||||||:|||||||||||| ||||::|:||||:||||||||| |||
Db    543 HAGIFTFEEPVTHVSESIGIMEVKVLRTSGARGNVIVPYKTIEGTARGGGEDFEDTCGEL 602

Qy    588 EFKNDE 593
          ||:|||
Db    603 EFQNDE 608


RESULT 10
ABG35474
ID   ABG35474 standard; Peptide; 609 AA.
XX
AC   ABG35474;
XX
DT   19-AUG-2002  (first entry)
XX
DE   Human peptide encoded by genome-derived single exon probe SEQ ID 25139.
XX
KW   Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease.
```

XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
[redacted] WPI; 2002-114183/15.

PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 27; SEQ ID No 25139; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 609 AA;

Query Match 66.4%; Score 2143.5; DB 23; Length 609;
Best Local Similarity 69.3%; Pred. No. 1.8e-209;

Matches 420; Conservative 76; Mismatches 89; Indels 21; Gaps 8;

```
Qy      1 MAWLRLQPLTSAFLHFGLVTFVLF--LNGLRAEAGGSGDVPSTGQNNESCSGSSDCKEGV 58
          |  | | |  |   |  :  :||  :: : ||    |:  ||:    |:||  ||:||
Db     11 MRRLSLSPTFSMGFHLLVTVSLLFSHVDHVIAETEMEGEGNETGE----CTGSYYCKKGV 66

Qy     59 ILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPN 118
          ||||| |::|| ||||||  |||||::|||||||||||||||:|||||||||:|:||||||
Db     67 ILPIWEPQDPSFGDKIARATVYFVAMVYMFLGVSIIADRFMSSIEVITSQEKEITIKKPN 126

Qy    119 GETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIVGSAAFNM 178
          |||: ||:|:|||||||||||||||||||||:|||||| | ||||||||||||||||||
Db    127 GETTKTTVRIWNETVSNLTLMALGSSAPEILLSVIEVCGHNFTAGDLGPSTIVGSAAFNM 186

Qy    179 FIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTL 238
          |||| :||||:|||||||||||||||:||||||||| |||:||:| |||||:|||||||
Db    187 FIIIALCVYVVPDGETRKIKHLRVFFVTAAWSIFAYTWLYIILSVISPGVVEVWEGLLTF 246

Qy    239 FFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKG---IEMDGKMMNSH 295
          ||||:||: |||||:|||||||::|:||  | ||:||| ||| |     |||||::|||
Db    247 FFFPICVVFAWVADRRLLFYKYVYKRYRAGKQRGMIIEHEGDRPSSKTEIEMDGKVVNSH 306

Qy    296 ---FLDGNLVPLEGKEVD----ESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQ 348
             |||| || ||  | |    |:|||| ||||:||||||:|:::||:|:|||  || ||
Db    307 VENFLDGALV-LEVDERDQDDEEARREMARILKELKQKHPDKEIEQLIELANYQVLSQQQ 365

Qy    349 KSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPE-DFISKVFFDPCSYQC 407
          |||||||||||:|||||||||:|||:||:|| || ||:|:  | | :||:||:  :|||
Db    366 KSRAFYRIQATRLMTGAGNILKRHAADQARKAVSMHEVNTEVTENDPVSKIFFEQGTYQC 425

Qy    408 LENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVG 467
          ||||| | ||:::|:|||:: |::||:::||||:||||||||||| |||:|||  ||
Db    426 LENCGTVALTIIRRGGDLTNTVFVDFRTEDGTANAGSDYEFTEGTVVFKPGDTQKEIRVG 485

Qy    468 IIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDD 527
          |||||||||||:| | ||||::  |  |:|:  |   |     | | ||  ||||| |||
Db    486 IIDDDIFEEDENFLVHLSNVKVSSEASEDGILEANHVS---TLACLGSPSTATVTIFDDD 542

Qy    528 HAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDTYGEL 587
          ||||||||    |||||||:|||||||||||||| ||||::|:||||:|||||||| |||
Db    543 HAGIFTFEEPVTHVSESIGIMEVKVLRTSGARGNVIVPYKTIEGTARGGGEDFEDTCGEL 602

Qy    588 EFKNDE 593
          ||:|||
Db    603 EFQNDE 608
```

RESULT 11
AAB41497
ID AAB41497 standard; Protein; 952 AA.
XX
AC AAB41497;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1261 polypeptide sequence SEQ ID NO:2522.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.

XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC75706.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 1809-1812; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 952 AA;

Query Match 60.7%; Score 1961; DB 21; Length 952;
Best Local Similarity 65.7%; Pred. No. 1.7e-190;
Matches 371; Conservative 87; Mismatches 83; Indels 24; Gaps 6;

```
Qy     40 STGQNNESCSGSSDCKEGVILPIWYPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFM 99
          |||     | ||  |: ||:||:| |::|||||| || :|||||::|||||||||||||
Db     70 STG----GCQGSYRCQPGVLLPVWEPDDPSLGDKAARAVVYFVAMVYMFLGVSIIADRFM 125

Qy    100 ASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHG 159
          |:||||||:|:|:|| | |||||  |:|:|||||||||||||||||||||:||||||
Db    126 AAIEVITSKEKEITITKANGETSVGTVRIWNETVSNLTLMALGSSAPEILLSVIEVCGHN 185

Qy    160 FIAGDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYM 219
          | ||:||| ||||||||||::| :|:|||| ||:||||||||||:||:|||||:||||:
Db    186 FQAGELGPGTIVGSAAFNMFVVIAVCIYVIPAGESRKIKHLRVFFVTASWSIFAYVWLYL 245

Qy    220 ILAVFSPGVVQVWEGLLTLFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEG 279
          |||||||||||||| |||| ||||||: ||:|||||||||::|:||||   |||| ||
Db    246 ILAVFSPGVVQVWEALLTLVFPVCVVFAWMADKRLLFYKYVYKRYRTDPRSGIIIGAEG 305

Qy    280 DHPKGIEMDGKMMNSHFLDGNL-----VPLEGKEVDESRREMIRILKDLKQKHPEKDLDQ 334
          | || ||:||  : :    | |      | | :|:| ||||:|:|||||||||:|||:|
Db    306 DPPKSIELDGTFVGAE-APGELGGLGPGPAEARELDASRREVIQILKDLKQKHPDKDLEQ 364
```

```
Qy    335 LVEMANYYALSHQQKSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDF 394
          || :|||||| |||||||||||||||:||||||:|::|||: :::|:       || :|
Db    365 LVGIANYYALLHQQKSRAFYRIQATRLMTGAGNVLRRHAADASRRAAPAEGAGEDE-DDG 423

Qy    395 ISKVFFDPCSYQCLENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVV 454
           |::||:|  | ||||||:|||:|  :||: : | ||||:|||||| ||:|||::|||:|
Db    424 ASRIFFEPSLYHCLENCGSVLLSVTCQGGEGNSTFYVDYRTEDGSAKAGSDYEYSEGTLV 483

Qy    455 LKPGETQKEFSVGIIDDDIFEEDEHFFVRLSNVRIEEEQ----PEEGMPPAIFNSLPLPR 510
          ||||||||  :|||||||||||||||||| |:|: : |    |: |           |:
Db    484 FKPGETQKELRIGIIDDDIFEEDEHFFVRLLNLRVGDAQGMFEPDGG---------GRPK 534

Qy    511 AVLASPCVATVTILDDDHAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVE 570
          | :| :|||||||||||||||:|:   :|||| :| ::|:|:|:|||||| :|:|||:
Db    535 GRLVAPLLATVTILDDDHAGIFSFQDRLLHVSECMGTVDVRVVRSSGARGTVRLPYRTVD 594

Qy    571 GTAKGGGEDFEDTYGELEFKNDETV 595
          |||:|||  :||  ||||| :|||:
Db    595 GTARGGGVHYEDACGELEFGDDETM 619
```

RESULT 12
ABB61721
ID ABB61721 standard; Protein; 950 AA.
XX
AC ABB61721;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11955.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL05824.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 11955; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 950 AA;

Query Match 41.5%; Score 1339.5; DB 22; Length 950;

```
  Best Local Similarity   49.5%;  Pred. No. 5.7e-127;
  Matches  275;  Conservative  100;  Mismatches  132;  Indels   48;  Gaps   12;

Qy      53 DCKEGVILPIWYPE-NPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQERE 111
           :| ||::||:| |: | |:||:: |  |||| |||:|:||||||||||:|| ||| ||
Db      99 ECSEGLVLPLWMPQRNISVGDRLVRGFVYFVLLIYLFVGVSIIADRFMAAIEAITSIERA 158

Qy     112 VTIKKPNGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEVCGHGFIAGDLGPSTIV 171
           | :| ||       :|:|||||:|||||||||||||||:||:    | :|||| |||
Db     159 VVVKGPNNTKQVMHVRIWNETVANLTLMALGSSAPEILLSVIEIYAKDFESGDLGPGTIV 218

Qy     172 GSAAFNMFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQV 231
           ||||:|:|:|| :|:  || || |:|:|||||:|| :|:|||:||::||:||:|||: |
Db     219 GSAAYNLFMIIAVCMIWIPAGEVRRIRHLRVFFVTALFSVFAYVWLWLILSVFTPGVILV 278

Qy     232 WEGLLTLFFFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGK- 290
           || ::|| |||: || |::|::||| |||| | || :| || ::  | |    :||| :
Db     279 WEAIVTLLFFPLTVLWAYIAERRLLVYKYMDKNYRVNK-RGTVVAGEHDQ---VEMDAEK 334

Qy     291 -----MMNSHFLDGNLVPLEGKEVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALS 345
                |: |    ||    : :  ||:||| | :| :|:||:|: ||:||  ||    |:
Db     335 GPKQPMVTS--ARGN----DAEAFDEARREYITLLTELRQKYPDADLEQLEMMAQEQVLA 388

Qy     346 HQQKSRAFYRIQATRMMTGAGNILKKHAAEQAKKASSMSEVHT------DEPEDFISKVF 399
               |||||||||||| | |:||:::|    | :  | ::||        ||  |   :::
Db     389 RSSKSRAFYRIQATRKMVGSGNLMRK---IQERAHSDLTEVKAQLHAGDDEEADDPIRMY 445

Qy     400 FDPCSYQCLENCGAVLLTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGE 459
           |:|  |  :|||   : |||: ||:|    |:|:|:||:|:|| |:   :| :   ||
Db     446 FEPGHYTVMENCGEFEVRVVRR-GDISTYASVEYETQDGTASAGTDFVGRKGLLSFPPGV 504

Qy     460 TQKEFSVGIIDDDIFEEDEHFFVRLSNVRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVA 519
            :: | : :||||:|||| |:|| |       | ||:              || | :|
Db     505 DEQRFRIEVIDDDVFEEDECFYIRLFN-------PSEGVK------------LAVPMIA 544

Qy     520 TVTILDDDHAGIFTFECDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGED 579
           || |||||||||| |      ::||:|  |:||:| |||||||||: |   ||    :|
Db     545 TVMILDDDHAGIFAFTDSVFEITESVGRFELKVMRYSGARGTVIVPYWTENDTAT-ESKD 603

Qy     580 FEDTYGELEFKNDET 594
           :|   ||| |:|:|:
Db     604 YEGARGELVFENNES 618


RESULT 13
AAY58044
ID   AAY58044 standard; Protein; 539 AA.
XX
AC   AAY58044;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Arabidopsis thaliana 11 transmembrane domain transporter MHX.
XX
KW   11 transmembrane domain transporter; MHX; metal cation-proton exchanger;
KW   zinc; magnesium; stress tolerance; transgenic plant.
XX
OS   Arabidopsis thaliana.
XX
PN   WO9961616-A2.
XX
PD   02-DEC-1999.
XX
PF   25-MAY-1999;   99WO-IL00277.
XX
PR   26-MAY-1998;   98IL-0124653.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (FLAN-) FLANDER INTERUNIVERSITY INST BIOTECHNOLO.
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Galili G,  Shaul O,  Inze D,  Van Montagu M,  Hilgemann DW;
XX
DR   WPI; 2000-062714/05.
DR   N-PSDB; AAZ47475, AAZ47476.
XX
PT   New isolated DNA encoding a metal cation-proton exchanger from plants,
PT   used to produce transgenic plants with improved stress tolerance  -
XX
PS   Claim 3; Fig 1; 52pp; English.
XX
CC   This is the Arabidopsis thaliana 11 transmembrane domain transporter
CC   designated MHX protein sequence. The MHX protein is a member of the 11-12
CC   transmembrane domain transporter family and has Mg2+/H+ or Zn2+/H+
CC   exchange activity. In plants Mg2+ and Zn2+ are essential for the
CC   integrity of ribosomes and cellular function require a fine balance of
CC   various ions including magnesium and zinc. The polypeptide and nucleotide
CC   sequences of MHX can be used to transform plant cells and produce
CC   transgenic plants. The transformed plants have improved tolerance to
CC   stress conditions such as drought, temperature, mineral excess or
CC   deficiency and high salinity. Other improvements include better growth
CC   on calcareous soils, possibly also improved uptake of important metals
CC   where these have limited availability in the soil.
XX
SQ   Sequence   539 AA;

  Query Match             11.2%;  Score 362.5;  DB 21;  Length 539;
  Best Local Similarity   31.3%;  Pred. No. 1.6e-27;
  Matches   91;  Conservative   61;  Mismatches   96;  Indels   43;  Gaps    9;

Qy      64 YPENPSLGDKIARVIVYFVALIYMFLGVSIIADRFMASIEVITSQEREVTIKKPNGETST 123
           :|   :| | : | ::||: | | |:|:| |  ||  |:| :    |:|    |  :
Db      29 FPGENTLSDGL-RGVLYFLGLAYCFIGLSAITARFFKSMENVVKHSRKVVTIDPITKAEV 87

Qy     124 TTI-RVWNETVSNLTLMALGSSAPEILLSLIEV---CGHGFIAGDLGPSTIVGSAAFNMF 179
           |  :||| |:::::|:| |:| |:| |: |:     |  : || ||| |:|||||::|
Db      88 ITYKKVWNFTIADISLLAFGTSFPQISLATIDAIRNMGERY-AGGLGPGTLVGSAAFDLF 146

Qy     180 IIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAYIWLYMILAVFSPGVVQVWEGLLTLF 239
           |  :|| |   || :||  | |: :   || :|||||:|| |:|| |: : | |||:
Db     147 PIHAVCVVVPKAGELKKISDLGVWLVELVWSFWAYIWLYIILEVWSPNVITLVEALLTVL 206

Qy     240 FFPVCVLLAWVADKRLLFYKYMHKKYRTDKHRGIIIETEGDHPKGIEMDGKMMNSHFLDG 299
           : : :: |:  |||   : |:              :  : || |:
Db     207 QYGLLLVHAYAQDKR---WPYLS----------LPMSRGDRPE---------------- 236

Qy     300 NLVPLEGKEVDESR----REMIRILKDLKQKHPEKDLDQLVEMANYYALSH 346
             ||   :|:| |:     ::  :  |  |   |     :|:: : :: ::
Db     237 EWVP---EEIDTSKDDNDNDVHDVYSDAAQDAVESGSRNIVDIFSIHSANN 284


RESULT 14
AAM78712
ID   AAM78712 standard; Protein; 618 AA.
XX
AC   AAM78712;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1374.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
```

PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK51845.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 3628-3629; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 618 AA;

Query Match 6.3%; Score 203.5; DB 22; Length 618;
Best Local Similarity 22.3%; Pred. No. 3.6e-11;
Matches 128; Conservative 91; Mismatches 204; Indels 151; Gaps 28;

```
Qy     39 PSTGQNNESCSGSSDCKEGVILPIWYPENP-SLGD-KIARVIVYFVALIYMFLGVSIIAD 96
          |   :  ||   |:|  :|      ||::  || : :   :|:: : :||||: ::|: |
Db     58 PPLSKEGES-ENSTDHAQGD-----YPKDIFSLEERRKGAIILHVIGMIYMFIALAIVCD 111

 y     97 R-FMASIEVITSQEREVTIKKPNGETSTTTIRVWNETVSNLTLMALGSSAPEILLSLIEV 155
          |: |: ||| :                : : :: |:  | || | ||||:  ||| |
Db    112 EFFVPSLTVITEK----------------LGI-SDDVAGATFMAAGGSAPELFTSLIGV 153

Qy    156 CGHGFIA-GDLGPSTIVGSAAFNMFIIIGICVYVIPDGETRKIKHLRVFFITAAWSIFAY 214
              |||  ::|  |||||| ||:  :||:|        :|:| :|        | :|
Db    154 ----FIAHSNVGIGTIVGSAVFNILFVIGMCALF-----SREILNL------TWWPLFRD 198

Qy    215 IWLY-----MILAVFSPGVVQVWEGLLTL-FFFPVCVLL-------AWVADKRLLFYKYM 261
          :  |     |::  |   |:  || || |  :|   | :         ||  |:::    :
Db    199 VSFYIVDLIMLIIFFLDNVIMWWESLLLLTAYFCYVVFMKFNVQVEKWV--KQMINRNKV 256

Qy    262 HKKYRTDKHRGIIIETEGDHP-----KGIEMDGKMMNSH--FLDGNLVPLEGKEVDESRR 314
           |   :        : | |       ::  |   : |   :  ::  |
Db    257 VKVTAPEAQAKPSAARDKDEPTLPAKPRLQRGGSSASLHNSLMRNSIFQL---------- 306

Qy    315 EMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQATRMMTGAGNILKKHAA 374
           ||  |  | :           |: :|  | :               ||  |   :| :
Db    307 -MIHTLDPLAE----------ELGSYGKLKYYD-----------TMTEEGRFREKASI 342

Qy    375 --EQAKKASSMSEVHTDEPE------DFISKVFFDPCSYQCLENCGAVLLTVVRKGGDMS 426
            : |||     : | || |      : : |:     :   :|   :   :   : |::|
Db    343 LHKIAKK-----KCHVDENERQNGAANHVEKIELPNSTSTDVEMTPSSDASEPVQNGNLS 397

Qy    427 KTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFEEDEHFFVRLSN 486
          :       :|:    || | :  :  | ||:|: :  |:   :|     :: | : :
Db    398 HNI-------EGAEAQTADEEEDQPLSLAWPSETRKQVTFLIVFPIVFP----LWITLPD 446

Qy    487 VRIEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDDDHAGIFTFECDTI----HVS 542
          ||  :    :  |   | |:                          |  |   :    |
Db    447 VR--KPSSRKFFPITFFGSI---------------------TWIAVFSYLMVWWAHQVG 482

Qy    543 ESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGG 576
          |:||: |  :  |  | || |    |    |: |
Db    483 ETIGISEEIMGLTILAAGTSIPDLITSVIVARKG 516
```

RESULT 15
ABB12262
ID ABB12262 standard; peptide; 546 AA.
XX
AC ABB12262;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human very large GPCR-1 homologue, SEQ ID NO:2632.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR N-PSDB; ABA09506.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
PS Claim 20; Page 321-322; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or

```
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a novel human
CC   polypeptide of the invention.
XX
SQ   Sequence   546 AA;

  Query Match             6.1%;  Score 197;  DB 22;  Length 546;
  Best Local Similarity   22.3%;  Pred. No. 1.3e-10;
  Matches   76;  Conservative   61;  Mismatches  128;  Indels   76;  Gaps   13;

Qy     299 GNLVPLEGK-EVDESRREMIRILKDLKQKHPEKDLDQLVEMANYYALSHQQKSRAFYRIQ 357
           |      ||  : : :  :| ::: |    ||:                  ::||  |
Db      64 GEFAETSGKLTMRDEQSAVIVVIQALNDDIPEE-------------------KSFYEFQ 103

Qy     358 ATRMMTGAGNILKKHAAEQAKKASSMSEVHTDEPEDFISKVFFDPCSYQCLENCGA--VL 415
           | : |          ::   :::: | :| |     :  |   |:: |    |  |
Db     104 LTAVSEGG-------VLSESSSTANITVVASDSP---YGRFAF---SHEQLRVSEAQRVN 150

Qy     416 LTVVRKGGDMSKTMYVDYKTEDGSANAGADYEFTEGTVVLKPGETQKEFSVGIIDDDIFE 475
           :|::|  ||    : : |||  |:| || |:    | :: : || :|   | |:|||  |
Db     151 ITIIRSSGDFGH-VRLWYKTMSGTAEAGLDFVPAAGELLFEAGEMRKSLHVEILDDDYPE 209

Qy     476 EDEHFFVRLSNVR---------IEEEQPEEGMPPAIFNSLPLPRAVLASPCVATVTILDD 526
             | | : :: |          |:|   :   || | |           :  : |: :
Db     210 GPEEFSLTITKVELQGRGYDFTIQENGLQIDQPPEIGNI-----------SIVRIIIMKN 258

Qy     527 DHA-GIFTFE--CDTIHVSESIGVMEVKVLRTSGARGTVIVPFRTVEGTAKGGGEDFEDT 583
           |:| ||  |:       | | :|:: : |:|  |  | |   | :   :|  || |:
Db     259 DNAEGIIEFDPKYTAFEVEEDVGLIMIPVVRLHGTYGYVTADFISQSSSASPGGVDYILH 318

Qy     584 YGELEFKNDE-------TVCDRQEADY---------GRRGG 608
             : |:: :       :: |  |:::          |  ||
Db     319 GSTVTFQHGQNLSFINISIIDDNESEFEEPIEILLTGATGG 359


Search completed: November 30, 2002, 12:28:08
Job time : 38.8079 secs
```